Multi-User MIMO Scheduling in the Fourth Generation Cellular Uplink

Narayan Prasad[1], Honghai Zhang[1], Hao Zhu [2], and Sampath Rangarajan[1]

[1]{prasad,honghai,sampath}@nec-labs.com; NEC Labs America

[2]zhuh@umn.edu; University of Minnesota

Abstract

In this paper, we consider Multi-User MIMO (MU-MIMO) scheduling in the 3GPP LTE-Advanced (3GPP LTE-A) cellular uplink. The 3GPP LTE-A cellular network is one of the two *true* fourth generation (4G) cellular networks as per the international telecommunications union and is expected to be the most widely deployed 4G cellular network. The 3GPP LTE-A uplink allows for precoded multi-stream (precoded MIMO) transmission from each scheduled user and also allows flexible multi-user (MU) scheduling wherein multiple users can be assigned the same time-frequency resource. However, exploiting these features is made challenging by certain practical constraints that have been imposed in order to maintain a low signaling overhead. We show that while the scheduling problem in the 3GPP LTE-A cellular uplink is NP-hard, it can be formulated as the maximization of a submodular set function subject to one matroid and multiple knapsack constraints. We then propose constant-factor polynomial-time approximation algorithms and demonstrate their superior performance via simulations. An interesting corollary that follows from our result is that a popular transmit antenna selection problem in point-to-point MIMO communications can be posed as a sub-modular maximization problem that is NP-hard but can be approximately solved (with at-least half optimality) by a simple greedy algorithm.

Keywords: Knapsack, Multi-user scheduling, Matroid, NP-hard, Resource allocation, Submodular maximization.

1 Introduction

The 3GPP LTE-A based cellular network [1] together with the IEEE 802.16m based cellular network are the only two cellular networks classified as fourth generation cellular networks by the international telecommunications union. Some key attributes that a 4G uplink must possess are the ability to support a peak spectral efficiency of 15 bps/Hz and a cell average spectral efficiency of 2 bps/Hz, ultra-low latency and bandwidths of up to 100MHz. To achieve these ambitious specifications, the 3GPP LTE-A uplink is based on a modified form of the orthogonal frequency-division multiplexing based multiple-access (OFDMA) [1]. In addition, it allows precoded multi-stream (precoded MIMO) transmission from each scheduled user as well as flexible multi-user scheduling. Notice that while OFDMA itself allows for significant spectral efficiency gains via channel dependent frequency domain scheduling, multi-user multi-stream communication promises substantially higher degrees of freedom [2, 3].

Our focus in this paper is on the 3GPP LTE-A uplink (UL) and in particular on MU MIMO scheduling for the LTE-A UL. Predominantly almost all of the 4G cellular systems that will be deployed will be based on the 3GPP LTE-A standard [1]. This standard is an enhancement of the basic LTE standard which is referred to in the industry as Release 8 [4] and indeed deployments conforming to Release 8 are already underway. The scheduling in the LTE-A UL is done in the frequency domain where in each scheduling interval the scheduler assigns one or more resource blocks (RBs) to each scheduled user. Each RB contains a pre-defined set of consecutive subcarriers and is the minimum allocation unit. The UL in the LTE-A network employs a modified form of OFDMA, referred to as the DFT-Spread-OFDMA . Here, each user employs a DFT precoder to spread its data symbols before placing them on its assigned RBs. In Fig. 1, we depict a feasible allocation in LTE-A UL MU scheduling. Notice that each user can be assigned up-to two mutually non-contiguous chunks, where each chunk is a set of contiguous RBs. This constraint together with the DFT spreading done by each user ensures that the peak to average power ratio of each user is kept in check. Notice also that there can be partial overlaps (in terms of assigned RBs) among co-scheduled users. Moreover, since the LTE-A base-station is expected to deploy advanced receivers, it is reasonable to assume that there is no explicit limit on the number of users that can be co-scheduled on an RB. The LTE-A UL also allows for precoded MIMO transmission from each scheduled user in order to achieve even higher data rates. While enabling multi-stream transmission can boost the user-rate, precoding confers the ability to steer the transmitted streams along suitable directions (in a signal space). In single-user

(SU) MIMO scheduling the suitable directions are the dominant eigen-directions of the scheduled user's channel, whereas in MU-MIMO scheduling the suitable directions also depend on the channels of the other overlapping co-scheduled users.

Some more practical constraints have been imposed on UL scheduling in 3GPP LTE-A. These include ones that seek to minimize the signaling overhead such as allowing each scheduled user to transmit with only one power level (or power spectral density (PSD)) on all its assigned RBs, [1] as well as enforcing that a scheduled user can be assigned no more than one precoding matrix in a scheduling interval. In addition, constraints that aim to mitigate intercell interference are also imposed along with those that arise due to the limited capacity of the downlink control channel on which the scheduling decisions are conveyed to the users.

The goal of this work is to design practical uplink MU-MIMO resource allocation algorithms for the LTE-A cellular network, where the term resource refers to RBs as well as precoding matrices. In particular, we consider the design of resource allocation algorithms via weighted sum rate utility maximization that account for finite user queues (buffers) and finite precoding codebooks. In addition, the designed algorithms comply with all the aforementioned practical constraints on the assignment of RBs and precoders to the scheduled users. Our main contributions are as follows:

1. We first assume that users can employ ideal Gaussian codes and that the base-station (BS) can employ an optimal receiver. We then enforce user rates to lie in a fundamental achievable rate region of the multiple access channel which is a polymatroid and show that the resulting resource allocation problem is NP-hard. We prove that the resource allocation problem can however be formulated as the maximization of a monotonic sub-modular set function subject to one matroid and multiple knapsack constraints, and can be solved using a recently discovered polynomial time randomized constant-factor approximation algorithm [5, 6]. We also adapt a simpler deterministic greedy algorithm and show that it yields a constant-factor approximation for several scenarios of interest.

2. We then consider practical scenarios where users employ codes constructed over finite alphabets. In this case the mutual information terms needed to specify an achievable rate region do not have closed form expressions. On the other hand the achievable rate region obtained for Gaussian alphabets can be a loose outer bound. Consequently, we obtain a tighter outer bound which is also a polymatroid.

[1] This PSD is implicitly determined by the number of RBs assigned to that user, i.e., the user divides its total power equally among all its assigned RBs subject possibly to a spectral mask constraint.

As a result all algorithms developed for Gaussian alphabets can be reused after simple modifications.

3. We provide simulation results using a realistic channel model and employ a data-dependent upper bound to benchmark the performance of our algorithms. We show that our LTE-A scheduling algorithm has a good average performance within $65 - 75\%$ of the upper bound.

1.1 Related Work

Resource allocation for the OFDMA networks has received significant attention [7–11] with most of it directed towards the downlink. A large fraction of the resource allocation problems hitherto considered are single-user (SU) scheduling problems, which attempt to maximize a system utility under the constraint that scheduled users can only be assigned non-overlapping subcarriers. These problems have been formulated as *continuous optimization problems*, and since they are in general non-linear and non-convex, many approaches including those based on game theory [12], dual decomposition [7] and the analysis of optimality conditions [13] have been developed. MU-MIMO in the downlink has been considered in [14, 15] where capacity scaling under imperfect channel estimation and/or quantized channel state information feedback is investigated but the design of approximation algorithms for resource allocation is not considered. Recent works have focused on emerging cellular standards and have formulated the resource allocation problems as constrained integer programs. Prominent examples are [10], [16] which consider the design of downlink SU-MIMO schedulers for LTE and LTE-A systems, respectively, and derive constant factor approximation algorithms. On the other hand, resource allocation for the DFT-Spread-OFDMA uplink has garnered relatively much less attention with [17–19] being the recent examples. In particular, [17, 18] show that the single-user UL LTE (Release 8) scheduling problem is NP-hard and provide constant-factor approximation algorithms, whereas [19] considers SU-MIMO LTE-A scheduling. The algorithms in [17–19] cannot incorporate MU scheduling and also cannot incorporate knapsack constraints. MU scheduling for the LTE (Release 8) UL is considered in detail in [20]. However, we emphasize that certain additional constraints imposed on LTE (Release 8) MU scheduling essentially ensure that algorithms optimized for LTE UL scheduling are unsuitable for LTE-A scheduling whereas algorithms optimized for LTE-A UL scheduling (as presented in this paper) are not even applicable to LTE UL scheduling since they yield infeasible solutions. To the best of our knowledge the design of approximation algorithms for MU-MIMO scheduling in LTE-A uplink has not been considered before.

2 MU-MIMO Scheduling in the LTE-A UL

Consider a single-cell with K users and one base-station (BS) which is assumed to have $N_r \geq 1$ receive antennas. Suppose that user k has $N_t \geq 1$ transmit antennas and its power budget is P_k. Let $\mathbf{H}_k^{(n)}$ denote the $N_r \times N_t$ channel matrix seen by the BS from user k on RB n. We let N denote the total number of RBs. For convenience and without loss of generality, in the following analysis we assume each RB to have unit size.

We consider the problem of scheduling users in the frequency domain in a given scheduling interval. Let $\alpha_k,\ 1 \leq k \leq K$ denote the non-negative weight of the k^{th} user which is an input to the scheduling algorithm and is updated using the output of the scheduling algorithm in every scheduling interval, say according to the proportional fairness rule [21]. Letting r_k denote the rate assigned to the k^{th} user (in bits per N RBs), we consider the following weighted sum rate utility maximization problem,

$$\max \sum_{1\leq k\leq K} \alpha_k r_k, \tag{1}$$

where the maximization is over the assignment of RBs, precoders and powers to the users **subject to:**

- **Decodability constraint:** The rates assigned to the scheduled users should be decodable by the base-station receiver. Notice that unlike SU-MIMO, MU-MIMO scheduling allows for multiple users to be assigned the same RB. As a result the rate that can be achieved for user k need not be only a function of the RBs, precoders and powers assigned to the k^{th} user but can also depend on those assigned to the other users as well.

- **One precoder and one power level per user:** Each scheduled user can be assigned any one precoding matrix from a finite codebook of such matrices $\mathcal{W}$. In addition, each scheduled user can transmit with only one power level (or power spectral density (PSD)) on all its assigned RBs. This PSD is implicitly determined by the number of RBs assigned to that user, i.e., the user divides its total power equally among all its assigned RBs.

- **At most two chunks per-user:** The set of RBs assigned to each scheduled user should form at-most two mutually non-contiguous chunks, where each chunk is a set of contiguous RBs. This constraint is a compromise between the need to provide enough scheduling flexibility and the need to keep the per-user peak-to-average-power ratio (PAPR) under check. A feasible RB allocation and

co-scheduling of users in LTE-A multi-user uplink is depicted in Fig 1.

- **Finite buffers** We let Q_k denote the size in bits of the queue (buffer) associated with the k^{th} user. Thus, the rate r_k assigned to user k cannot exceed Q_k.

- **Control channel overhead and interference limit constraints:** Every user that is scheduled on at least one RB must be informed about its transmission rate and the set of RBs on which it must transmit along with the precoder it should employ. This information is sent on the DL control channel of limited capacity which in turn imposes a limit on the set of users that can be scheduled. On the other hand, the scheduling decisions that are made must respect limits imposed to mitigate the interference caused to other cells. In [20] it is shown that the control channel overhead constraints can be modeled as binary column-sparse knapsack constraints, whereas the interference limit constraints can be modeled as generic knapsack constraints.

We will formulate the optimization problem in (1) as the *maximization of a monotonic submodular set function subject to one matroid and multiple knapsack constraints.*

Towards this end, let $\underline{e} = (u, \mathbf{c}, \mathbf{W})$ denote an element, where $1 \leq u \leq K$ denotes a user, $\mathbf{W} \in \mathcal{W}$ denotes a precoder from a finite codebook $\mathcal{W}$ and $\mathbf{c} \in \boldsymbol{\mathcal{C}}$ denotes a valid assignment of RBs chosen from the set $\boldsymbol{\mathcal{C}}$ containing all possible valid assignments. In particular, each $\mathbf{c}$ is a vector with binary-valued ($\{0, 1\}$) elements and we say an RB i belongs to $\mathbf{c}$ ($i \in \mathbf{c}$) if $\mathbf{c}$ contains a one in its i^{th} position, i.e., $c(i) = 1$. Note that the non-zero entries in each $\mathbf{c} \in \boldsymbol{\mathcal{C}}$ form at-most two non-contiguous chunks. Next, we let $\underline{\mathcal{E}} = \{\underline{e} = (u, \mathbf{c}, \mathbf{W}) : 1 \leq u \leq K, \mathbf{c} \in \boldsymbol{\mathcal{C}}, \mathbf{W} \in \mathcal{W}\}$ denote the ground set of all possible such elements. For any such element we adopt the convention that

$$\begin{aligned} \underline{e} = (u, \mathbf{c}, \mathbf{W}) \Rightarrow \mathbf{c}_{\underline{e}} = \mathbf{c}; \;\; \mathbf{W}_{\underline{e}} = \mathbf{W}; \;\; u_{\underline{e}} = u; \\ \alpha_{\underline{e}} = \alpha_u; \; Q_{\underline{e}} = Q_u; \; \mathbf{H}_{\underline{e}}^{(n)} = \mathbf{H}_u^{(n)} \; \forall \, n. \end{aligned} \tag{2}$$

In addition, we let $p_{\underline{e}}$ denote the power level (PSD) associated with the element $\underline{e} = (u, \mathbf{c}, \mathbf{W})$. This PSD can be computed as $\frac{P_u}{\text{size}(\mathbf{c})}$, where size($\mathbf{c}$) denotes the number of ones (number of RBs) in $\mathbf{c}$. Let $\alpha_{\underline{e}}, Q_{\underline{e}}$ denote the weight and buffer (queue) size associated with the element $\underline{e}$, respectively and let $r_{\underline{e}}$ denote the rate associated with the element $\underline{e}$. We will use the phrase *selecting an element* $\underline{e}$ *to imply that the user* $u_{\underline{e}}$ *is scheduled to transmit on the RBs indicated in* $\mathbf{c}_{\underline{e}}$ *with PSD* $p_{\underline{e}}$ *and precoder* $\mathbf{W}_{\underline{e}}$. Thus, *the constraints*

of one precoder and one power level per user along with at most two chunks per-user can be imposed by allowing the scheduler to select any subset of elements $\underline{\mathcal{U}} \subseteq \underline{\mathcal{E}}$ such that $\sum_{\underline{e}\in\underline{\mathcal{U}}} 1\{u_{\underline{e}} = u\} \leq 1$ for each $u \in \{1, \cdots, K\}$, where $1\{.\}$ denotes the indicator function. Accordingly, we define a family of subsets of $\underline{\mathcal{E}}$, denoted by $\underline{\mathcal{I}}$, as

$$\underline{\mathcal{I}} = \left\{ \underline{\mathcal{U}} \subseteq \underline{\mathcal{E}} : \sum_{\underline{e}\in\underline{\mathcal{U}}} 1\{u_{\underline{e}} = u\} \leq 1, \;\; \forall \;\; 1 \leq u \leq K \right\}. \tag{3}$$

We next consider the decodability constraint after first assuming that each user can employ ideal Gaussian codes (i.e., codes for which the coded modulated symbols can be regarded as i.i.d. Gaussian) and that the BS can employ an optimal receiver. Subsequently, we will consider finite input alphabets. Recall that in DFT-Spread-OFDMA each user linearly transforms its codeword using a DFT matrix in order to reduce the PAPR. Note, however, that under the assumption of ideal Gaussian codes the DFT spreading operation performed by each user on its codeword has no effect. This is because i.i.d. Gaussian distribution is invariant with respect to any unitary linear transformation. Accordingly, we define a set function $f : 2^{\underline{\mathcal{E}}} \to I\!R_+$ as

$$f(\underline{\mathcal{U}}) = \sum_{n=1}^{N} \log \left| \mathbf{I} + \sum_{\underline{e}\in\underline{\mathcal{U}}} p_{\underline{e}} c_{\underline{e}}(n) \mathbf{H}_{\underline{e}}^{(n)} \mathbf{W}_{\underline{e}} (\mathbf{H}_{\underline{e}}^{(n)} \mathbf{W}_{\underline{e}})^{\dagger} \right|, \tag{4}$$

for all $\underline{\mathcal{U}} \subseteq \underline{\mathcal{E}}$. It can be verified that $f(.)$ defined in (4) is a submodular set function (see for instance Proposition 2 in [22]), i.e.,

$$f(\underline{\mathcal{A}} \cup \{\underline{e}\}) - f(\underline{\mathcal{A}}) \geq f(\underline{\mathcal{B}} \cup \{\underline{e}\}) - f(\underline{\mathcal{B}}),$$

for all $\underline{\mathcal{A}} \subseteq \underline{\mathcal{B}} \subseteq \underline{\mathcal{E}}$ and $\underline{e} \in \underline{\mathcal{E}}$. Further since it is monotonic (i.e., $f(\underline{\mathcal{A}}) \leq f(\underline{\mathcal{B}})$, $\forall\, \underline{\mathcal{A}} \subseteq \underline{\mathcal{B}}$) and normalized $f(\phi) = 0$, where ϕ denotes the empty set, we can assert that $f(.)$ is a rank function. Consequently, for each $\underline{\mathcal{U}} \subseteq \underline{\mathcal{E}}$, the region

$$\underline{\mathcal{P}}(\underline{\mathcal{U}}, f) = \left\{ \mathbf{r} = [r_{\underline{e}}]_{\underline{e}\in\underline{\mathcal{U}}} \in I\!R_+^{|\underline{\mathcal{U}}|} : \sum_{\underline{e}\in\underline{\mathcal{A}}} r_{\underline{e}} \leq f(\underline{\mathcal{A}}), \; \forall\; \underline{\mathcal{A}} \subseteq \underline{\mathcal{U}} \right\}, \tag{5}$$

is a polymatroid [3]. Note that for each $\underline{\mathcal{U}} \subseteq \underline{\mathcal{E}}$, $\underline{\mathcal{P}}(\underline{\mathcal{U}}, f)$ is the fundamental achievable rate region of a

multiple access channel so that each rate-tuple $\mathbf{r}_{\underline{\mathcal{U}}} = [r_{\underline{e}}]_{\underline{e}\in\underline{\mathcal{U}}} \in \underline{\boldsymbol{\mathcal{P}}}(\underline{\mathcal{U}}, f)$ is achievable [3, 23]. In particular, each rate-tuple $\mathbf{r}_{\underline{\mathcal{U}}} = [r_{\underline{e}}]_{\underline{e}\in\underline{\mathcal{U}}} \in \underline{\boldsymbol{\mathcal{P}}}(\underline{\mathcal{U}}, f)$ is achievable [3] in the sense that for any rate assignment arbitrarily close to $\mathbf{r}_{\underline{\mathcal{U}}}$ (i.e., $\mathbf{r} : \mathbf{r} < \mathbf{r}_{\underline{\mathcal{U}}}$) there exist coding and decoding schemes that can meet any acceptable level of error probability. Thus, *we can impose decodability constraints by imposing that the assigned rate-tuple satisfy* $\mathbf{r}_{\underline{\mathcal{U}}} \in \underline{\boldsymbol{\mathcal{P}}}(\underline{\mathcal{U}}, f)$ *for any selected subset* $\underline{\mathcal{U}} \subseteq \underline{\mathcal{E}}$.

Next, in order to impose buffer (queue) constraints, we define

$$\underline{\boldsymbol{\mathcal{B}}}(\underline{\mathcal{U}}) = \{\mathbf{r} = [r_{\underline{e}}]_{\underline{e}\in\underline{\mathcal{U}}} \in \mathbb{R}_+^{|\underline{\mathcal{U}}|} : 0 \le r_{\underline{e}} \le Q_{\underline{e}},\ \forall\, \underline{e} \in \underline{\mathcal{U}}\},\ \forall\, \underline{\mathcal{U}} \subseteq \underline{\mathcal{E}}. \tag{6}$$

Thus, for a (tentative) choice $\underline{\mathcal{U}}$*, we can satisfy both decodability and buffer constraints by assigning only rate-tuples that lie in the region* $\underline{\boldsymbol{\mathcal{P}}}(\underline{\mathcal{U}}, f) \cap \underline{\boldsymbol{\mathcal{B}}}(\underline{\mathcal{U}})$. Clearly among all such rate-tuples we are interested in the one that maximizes the weighted sum rate. Hence, without loss of optimality with respect to (1), with each $\underline{\mathcal{U}} \subseteq \underline{\mathcal{E}}$ we can associate a rate-tuple in $\underline{\boldsymbol{\mathcal{P}}}(\underline{\mathcal{U}}, f) \cap \underline{\boldsymbol{\mathcal{B}}}(\underline{\mathcal{U}})$ that maximizes the weighted sum rate. Consequently, we define the following set function that determines the reward obtained upon selecting any subset of $\underline{\mathcal{E}}$. We define the set function $h : 2^{\underline{\mathcal{E}}} \to \mathbb{R}_+$ as

$$h(\underline{\mathcal{U}}) = \max_{\substack{\mathbf{r}=[r_{\underline{e}}]_{\underline{e}\in\underline{\mathcal{U}}} \\ \mathbf{r}\in\underline{\boldsymbol{\mathcal{P}}}(\underline{\mathcal{U}},f)\cap\underline{\boldsymbol{\mathcal{B}}}(\underline{\mathcal{U}})}} \left\{ \sum_{\underline{e}\in\underline{\mathcal{U}}} \alpha_{\underline{e}} r_{\underline{e}} \right\},\ \ \forall\, \underline{\mathcal{U}} \subseteq \underline{\mathcal{E}}. \tag{7}$$

Leveraging the arguments made in [20], we can represent the control channel overhead constraints as L packing (knapsack) constraints such that a subset $\underline{\mathcal{U}}$ is feasible if and only if

$$\mathbf{A}_C \mathbf{x}_{\underline{\mathcal{U}}} \le \mathbf{1}_L, \tag{8}$$

where $\mathbf{A}_C \in \{0,1\}^{L\times|\underline{\mathcal{E}}|}$ is a binary valued matrix and $\mathbf{1}_L$ is a L length vector of ones. Moreover, the total number of non-zero entries in any column of $\mathbf{A}_C$ is no more than an integer $\Delta \ge 1$ which denotes the column sparsity level. On the other hand, the interference limit constraints can be represented as

$$\mathbf{A}_I \mathbf{x}_{\underline{\mathcal{U}}} \le \mathbf{1}_M, \tag{9}$$

where $\mathbf{A}_I \in [0,1]^{M\times|\underline{\mathcal{E}}|}$ and $\mathbf{1}_M$ is a M length vector of ones.

Summarizing the aforementioned results, we have formulated (1) as the following optimization problem:

$$\begin{aligned} &\max_{\underline{\mathcal{U}}\subseteq\underline{\mathcal{E}}}\{h(\underline{\mathcal{U}})\} \text{ s.t.} \\ &\underline{\mathcal{U}}\in\underline{\mathcal{I}}; \\ &\mathbf{A}_I\mathbf{x}_{\underline{\mathcal{U}}}\leq\mathbf{1}_M;\ \ \mathbf{A}_C\mathbf{x}_{\underline{\mathcal{U}}}\leq\mathbf{1}_L. \end{aligned} \tag{10}$$

In (10) we regard M, Δ as constants that are arbitrarily fixed, whereas L can scale polynomially in the cardinality of the ground set $|\underline{\mathcal{E}}|$. Then, for a given number of users K, number of RBs N and the codebook cardinality $|\mathcal{W}|$ (which together fix $|\underline{\mathcal{E}}|$), an instance (or input) of the problem in (10) consists of a set of non-negative user weights $\{\alpha_u\}$ and queue sizes $\{Q_u\}$, per-user per-RB channel matrices $\{\mathbf{H}_u^{(n)}\}:\ 1\leq u\leq K, 1\leq n\leq N$, a codebook $\mathcal{W}$ (of cardinality $|\mathcal{W}|$) along with a column sparse matrix $\mathbf{A}_C\in\{0,1\}^{L\times|\underline{\mathcal{E}}|}$ and any matrix $\mathbf{A}_I\in[0,1]^{M\times|\underline{\mathcal{E}}|}$. The output is a subset $\hat{\underline{\mathcal{U}}}\subseteq\underline{\mathcal{E}}$ along with a rate-tuple $r_{\hat{\underline{\mathcal{U}}}}$. Note that $|\underline{\mathcal{E}}|$ is $O(K|\mathcal{W}|N^4)$.

We first introduce the following two results that will be invoked later.

Lemma 1. *The family of subsets $\underline{\boldsymbol{\mathcal{I}}}$ defined in (3) is an independence family and $(\underline{\mathcal{E}},\underline{\boldsymbol{\mathcal{I}}})$ is a partition matroid.*

Proof. First we note that $\underline{\boldsymbol{\mathcal{I}}}$ is downward closed, i.e., if $\underline{\mathcal{A}}\in\underline{\boldsymbol{\mathcal{I}}}$ then any $\underline{\mathcal{B}}\subseteq\underline{\mathcal{A}}$ satisfies $\underline{\mathcal{B}}\in\underline{\boldsymbol{\mathcal{I}}}$. Next, let $\underline{\mathcal{E}}_{(k)}$ denote the set of all $\underline{e}\in\underline{\mathcal{E}}: u_{\underline{e}}=k$ and notice that $\underline{\mathcal{E}}_{(k)}\cap\underline{\mathcal{E}}_{(j)}=\phi,\ \forall\ k\neq j$. Then, note that $\underline{\boldsymbol{\mathcal{I}}}$ can also be defined as $\underline{\mathcal{A}}\in\underline{\boldsymbol{\mathcal{I}}}\Leftrightarrow|\underline{\mathcal{A}}\cap\underline{\mathcal{E}}_{(k)}|\leq 1\ \forall\ 1\leq k\leq K$. Further, it can be verified $\underline{\boldsymbol{\mathcal{I}}}$ satisfies the exchange property, i.e., for any $\underline{\mathcal{A}},\underline{\mathcal{B}}\in\underline{\boldsymbol{\mathcal{I}}}$ such that $|\underline{\mathcal{A}}|>|\underline{\mathcal{B}}|$ we have that $\exists\ \underline{e}\in\underline{\mathcal{A}}\setminus\underline{\mathcal{B}}$ such that $\underline{\mathcal{B}}\cup\{\underline{e}\}\in\underline{\boldsymbol{\mathcal{I}}}$. Thus, we can conclude that $(\underline{\mathcal{E}},\underline{\boldsymbol{\mathcal{I}}})$ is a partition matroid. ◻

The proof of the following lemma follows from basic definitions [24] and is skipped for brevity.

Lemma 2. *The region $\underline{\boldsymbol{\mathcal{P}}}(\underline{\mathcal{U}},f)\cap\underline{\boldsymbol{\mathcal{B}}}(\underline{\mathcal{U}}),\ \forall\ \underline{\mathcal{U}}\subseteq\underline{\mathcal{E}}$ is a polymatroid characterized by the rank function $f':2^{\underline{\mathcal{E}}}\to\mathbb{R}_+$ where*

$$f'(\underline{\mathcal{U}})=\min_{\underline{\mathcal{R}}\subseteq\underline{\mathcal{U}}}\left\{f(\underline{\mathcal{U}}\setminus\underline{\mathcal{R}})+\sum_{\underline{e}\in\underline{\mathcal{R}}}Q_{\underline{e}}\right\},\ \forall\,\underline{\mathcal{U}}\subseteq\underline{\mathcal{E}}. \tag{11}$$

We are now ready to offer our main result. Let us assume that computing $h(\underline{\mathcal{U}})$ for any $\underline{\mathcal{U}}\subseteq\underline{\mathcal{E}}$ incurs a unit cost (or equivalently is given by an oracle in a single query). We will show that even under this

assumption the problem in (10) is NP hard.

Theorem 1. *The optimization problem in (10) is NP hard and is the maximization of a monotonic sub-modular set function subject to one matroid and multiple knapsack constraints.*

Proof. We will first show that (10) is the maximization of a monotonic sub-modular set function subject to one matroid and multiple knapsack constraints. Invoking Lemma 1, it suffices to show that the function $h(.)$ is a monotonic submodular set function. From the definition of $h(.)$ in (7) it is readily seen that it is monotonic, i.e., $h(\underline{\mathcal{U}}') \leq h(\underline{\mathcal{U}}),\ \forall\, \underline{\mathcal{U}}' \subseteq \underline{\mathcal{U}} \subseteq \underline{\mathcal{E}}$. Let $o(.,.)$ denote an ordering function such that for any subset $\underline{\mathcal{U}} \subseteq \underline{\mathcal{E}}$, $o(\underline{\mathcal{U}}, k)$ is the element having the k^{th} largest weight among the elements in $\underline{\mathcal{U}}$. Hence we have that $\alpha_{o(\underline{\mathcal{U}},1)} \geq \alpha_{o(\underline{\mathcal{U}},2)} \geq \alpha_{o(\underline{\mathcal{U}},|\underline{\mathcal{U}}|)}$. Further, let us adopt the convention that for any subset $\underline{\mathcal{U}} \subseteq \underline{\mathcal{E}}$, $o(\underline{\mathcal{U}}, k) = \phi,\ \forall\, k \geq |\underline{\mathcal{U}}| + 1$ & $\alpha_\phi = 0$. We can now invoke Lemma 2 together with the important property that the *rate-tuple in any polymatroid that maximizes the weighted sum is determined by the corner point of that polymatroid in which the elements are arranged in the non-increasing order of their weights [3, 24].* Thus, we can express $h(.)$ as

$$h(\underline{\mathcal{U}}) = \sum_{k=1}^{|\underline{\mathcal{U}}|} (\alpha_{o(\underline{\mathcal{U}},k)} - \alpha_{o(\underline{\mathcal{U}},k+1)}) f'(\{o(\underline{\mathcal{U}}, 1), \cdots, o(\underline{\mathcal{U}}, k)\}),\ \ \forall\, \underline{\mathcal{U}} \subseteq \underline{\mathcal{E}}. \tag{12}$$

A key step is to express (12) as

$$h(\underline{\mathcal{U}}) = \sum_{k=1}^{|\underline{\mathcal{E}}|} (\alpha_{o(\underline{\mathcal{E}},k)} - \alpha_{o(\underline{\mathcal{E}},k+1)}) \underbrace{f'(\{o(\underline{\mathcal{E}}, 1), \cdots, o(\underline{\mathcal{E}}, k)\} \cap \underline{\mathcal{U}})}_{f'_k(\underline{\mathcal{U}})},\ \ \forall\, \underline{\mathcal{U}} \subseteq \underline{\mathcal{E}}. \tag{13}$$

Note that in (13) the number of terms in the summation as well as the non-negative combining weights $\{(\alpha_{o(\underline{\mathcal{E}},k)} - \alpha_{o(\underline{\mathcal{E}},k+1)})\}$ do not depend on $\underline{\mathcal{U}}$. It can then be verified that since $f'(.)$ is monotonic and submodular, each set function $f'_k(.)$ is also a monotonic and submodular set function. From (13) it can now be inferred that since $h(.)$ is a weighted sum of monotonic submodular functions in which all the combining weights are non-negative, *it is a monotonic submodular set function.* Thus, (10) is the maximization of a monotonic submodular set function subject to one matroid and multiple knapsack constraints.

We will now show that (10) is an NP hard problem. We will consider instances of the problem where the number of RBs $N = 1$, all users have identical weights, unit powers, infinite queues and one transmit antenna each and where the codebook $\mathcal{W}$ is degenerate, i.e., $\mathcal{W} = \{1\}$. Thus, we have $|\underline{\mathcal{E}}| = K$. In addition,

we assume that the number of receive antennas is equal to the number of users K so that a given input of user channels forms a $K \times K$ matrix, denoted here by $\mathbf{H} = [\mathbf{h}_1, \cdots, \mathbf{h}_K]$. Further, we will assume only one knapsack constraint which in particular is a cardinality constraint on the number of users that can be scheduled on the one available RB. We will show that the problem specialized to these instances is also NP-hard so that the original problem is NP-hard. Note that the matroid constraint now becomes redundant and (10) simplifies to maximizing the sum rate under a cardinality constraint

$$\max_{\substack{\mathbf{D}=\mathrm{diag}\{d_1,\cdots,d_K\} \\ d_k\in\{0,1\}\ \forall\ k\ \&\ \sum_{k=1}^K d_k \le C}} \log|\mathbf{I} + \mathbf{H}\mathbf{D}\mathbf{H}^\dagger|, \tag{14}$$

where $C : 1 \le C \le K$ is the input maximum cardinality. Now using the determinant equality

$$\log|\mathbf{I} + \mathbf{H}\mathbf{D}\mathbf{H}^\dagger| = \log|\mathbf{I} + \mathbf{D}\mathbf{H}^\dagger\mathbf{H}\mathbf{D}| \tag{15}$$

together with the monotonicity of the objective function, we can re-write (14) as

$$\max_{\substack{\mathbf{D}=\mathrm{diag}\{d_1,\cdots,d_K\} \\ d_k\in\{0,1\}\ \forall\ k\ \&\ \sum_{k=1}^K d_k = C}} \log|\mathbf{I} + \mathbf{D}\mathbf{H}^\dagger\mathbf{H}\mathbf{D}|. \tag{16}$$

Note that (16) is equivalent to determining the $C \times C$ principal sub-matrix of the positive definite matrix $\mathbf{I} + \mathbf{H}^\dagger\mathbf{H}$ having the maximum determinant. Note that for a given K, an instance of the problem in (16) is the matrix $\mathbf{H}$ together with C. We will prove that (16) is NP-hard via contradiction. Suppose now that an efficient algorithm (with a complexity polynomial in K) exists that can optimally solve (16) for any input $K \times K$ matrix $\mathbf{H}$ and any $C : 1 \le C \le K$. This in turn would imply that there exists an efficient algorithm (with a complexity polynomial in K) that for any input $C : 1 \le C \le K$ and any $K \times K$ positive definite matrix Σ, can determine the $C \times C$ principal sub-matrix of Σ having the maximum determinant. Invoking the reduction developed in [25], this would then contradict the NP hardness of the problem of determining whether a given input graph has a clique of a given input size. $\square$

Theorem 2. *There is a randomized algorithm whose complexity scales polynomially in $|\underline{\mathcal{E}}|$ and which yields a $\frac{e-1}{e^2(M+\Delta+1)+o(M)}$ approximation to (10).*

Proof. The key observation is that the partition matroid constraint in (10) can be expressed as K linear packing constraints (one for each user). Let $\mathbf{A}_P$ denote the resulting $K \times |\underline{\mathcal{E}}|$ packing matrix whose k^{th}

row corresponds to the k^{th} user. Note that this row has ones in each position for which the corresponding element $\underline{e}$ satisfies $u_{\underline{e}} = k$ and zeros elsewhere. Together these K packing constraints are sparse packing constraints wherein in each column a non-zero entry appears only once. Thus, the total $K+L+M$ packing constraints are sparse constraints in which each element can appear in at-most $M+\Delta+1$ constraints so that each column can have at-most $M+\Delta+1$ non-zero entries. With this understanding, we can invoke the randomized algorithm from [6] which is applicable to the maximization of any monotonic submodular function subject to sparse packing constraints and obtain the guarantee claimed in the theorem. $\square$

Notice that since any monotonic submodular set function is also monotonic and sub-additive, we can infer the following result from Theorem 1.

Lemma 3. *The function $h(.)$ defined in (7) is sub-additive, i.e.,*

$$h(\underline{\mathcal{U}}) \leq h(\underline{\mathcal{U}}_1) + h(\underline{\mathcal{U}}_2),\ \forall\, \underline{\mathcal{U}}_1, \underline{\mathcal{U}}_2, \underline{\mathcal{U}} : \underline{\mathcal{U}}_1 \cup \underline{\mathcal{U}}_2 = \underline{\mathcal{U}}. \tag{17}$$

Practical implementation might demand a simpler and combinatorial (deterministic) algorithm. Unfortunately, as remarked in [5], it is difficult to design combinatorial (deterministic) algorithms that can combine both matroid and knapsack constraints. Nevertheless in Algorithm I we specialize a well known greedy algorithm to our problem of interest (10). Before analyzing the performance of Algorithm I we consider the following scenarios that involve simpler modeling of the constraints and are of particular interest. We first note that necessary and sufficient conditions for a knapsack constraint (with rational valued coefficients) to be a matroid constraint have been derived in [26]. A simple sufficient condition for a knapsack constraint to be matroid constraint is the following.

Lemma 4. *The i^{th} knapsack constraint is a matroid constraint if all its strictly positive coefficients are identical,i.e., $1\{A_{i,j} > 0\} = 1\{A_{i,k} > 0\} \Rightarrow A_{i,j} = A_{i,k},\ \forall\, j, k$.*

Then consider the scenarios that are covered by the following assumptions.

Assumption 1. *The control channel overhead constraints are modeled using L knapsack constraints but where L now represents the number of orthogonal (non-overlapping) control channel regions. Each user (and hence all its corresponding elements) is associated with only one of these regions. Further, each constraint corresponds to a cardinality constraint which enforces that no more than a given number of elements among those associated with the corresponding control region can be scheduled. Notice that these*

L control channel overhead constraints are sparse with $\Delta = 1$ and since they satisfy Lemma 4 they are matroid constraints as well.

Assumption 2. *For each adjacent victim BS, the elements of $\underline{\mathcal{E}}$ are divided into two sets using an appropriate threshold: one set comprising those which cause high interference and the other one comprising those which do not. Then a cardinality constraint is imposed only on the set of elements that cause high interference. Thus, all resulting interference limit constraints (upon considering M victim BSs) satisfy Lemma 4 and hence are matroid constraints.*

The following result provides the worst-case guarantee offered by Algorithm I.

Theorem 3. *The complexity of Algorithm I is $O(K^2N^4|\mathcal{W}|)$ and it yields a $\frac{1}{K}$ approximation to (10). Further, if Assumptions 1 and 2 are satisfied then Algorithm I yields a constant-factor $\frac{1}{2+M}$ approximation to (10).*

Proof. We first consider the complexity of Algorithm I and note that since the partition matroid constraint needs to be satisfied, there can be at-most K steps in repeat-until loop of the algorithm. Also, recall that the the size of the ground set $\underline{\mathcal{E}}$ is $O(KN^4|\mathcal{W}|)$. Then, at each step we need to compute $h(\underline{\mathcal{S}} \cup \underline{e})$ for each $\underline{e} \in \underline{\mathcal{E}} \setminus \underline{\mathcal{S}}$ such that $\underline{\mathcal{S}} \cup \underline{e}$ satisfies all the constraints. Thus, the worst-case complexity is $O(K^2N^4|\mathcal{W}|)$.

Let us now consider the approximation guarantees. Notice that due to the partition matroid constraint any optimal solution to (10) cannot contain more that K elements. Then, using the subadditivity of $h(.)$ shown in Lemma 3 together with the facts that Algorithm I is monotonic and in its first step selects the element of $\underline{\mathcal{E}}$ having the highest weighted rate, suffice to prove the $\frac{1}{K}$ guarantee. On the other hand, suppose that Assumptions 1 and 2 are satisfied (over all instances). Consider the L control channel constraints and let $\underline{\mathcal{E}}_\ell$ denote the set of elements involved in the ℓ^{th} control channel constraint so that $\underline{\mathcal{E}} = \cup_{\ell=1}^{L}\underline{\mathcal{E}}_\ell$. Recall that $\underline{\mathcal{E}}_\ell \cap \underline{\mathcal{E}}_{\ell'} = \phi,\ \ell \neq \ell'$ and notice that any set $\underline{\mathcal{U}} \subseteq \underline{\mathcal{E}}$ that satisfies these L constraints can be expressed as $\underline{\mathcal{U}} = \cup_{\ell=1}^{L}\underline{\mathcal{U}}_\ell$, where $\underline{\mathcal{U}}_\ell \subseteq \underline{\mathcal{E}}_\ell : |\underline{\mathcal{U}}_\ell| \leq C_\ell,\ 1 \leq \ell \leq L$, where C_ℓ is the cardinality bound imposed by the ℓ^{th} control channel constraint. Thus the L control channel constraints together are indeed one partition matroid. More importantly, the intersection of this partition matroid with the one defined in Lemma 1 is also one matroid. This can be verified by observing that all maximal members in this intersection have the same cardinality of $\min\{K, \sum_{\ell=1}^{L} C_\ell\}$. Finally, combining this matroid with the other M (interference limit) matroid constraints, we see that the feasible subsets belong to the intersection of $M+1$ matroids and

hence form a $p-$system where $p = M + 1$. Then invoking the guarantee offered by the greedy algorithm on a $p-$system [27, 28], proves the second part. $\square$

Remark 1. *Let us reconsider the submodular maximization problem defined in (14). This problem in fact also represents a popular transmit antenna selection problem in point-to-point MIMO communications [29]. Indeed, K can be regarded as the total number of available transmit antennas while C then denotes the number of transmit antennas that have to be selected and a normalization factor $\sqrt{\frac{\rho}{C}}$, where ρ denotes the SNR, can be absorbed into the matrix $\mathbf{H}$. Then, our result in Theorem 1 proves that this transmit antenna selection problem is NP-hard. Next, the greedy Algorithm I when specialized to this problem reduces to a known incremental successive transmit antenna selection algorithm [29] but for which no approximation guarantees were hitherto known. Notice that this problem satisfies Assumptions 1 and 2 since the constraint in (14) can be accommodated using just one control channel knapsack constraint that has equal coefficients for all users. Then, invoking the result in Theorem 3 (with $M = 0$) we can infer that the greedy Algorithm I (or equivalently the incremental successive transmit antenna selection algorithm) offers a $1/2$ approximation to the transmit antenna selection problem.*

Recall that hitherto we have assumed that computing $h(\underline{\mathcal{U}})$ for any $\underline{\mathcal{U}} \subseteq \underline{\mathcal{E}}$ incurs a unit cost. We can indeed show that Algorithm I has polynomial complexity under a stricter notion that computing $f(\underline{\mathcal{U}})$ (instead of $h(\underline{\mathcal{U}})$) for any $\underline{\mathcal{U}} \subseteq \underline{\mathcal{E}}$ incurs a unit cost.[2] To show this, it suffices to prove that $h(\underline{\mathcal{U}})$ can be determined with a complexity polynomial in $|\underline{\mathcal{U}}|$. A key observation towards this end is that for any $\underline{\mathcal{U}} \subseteq \underline{\mathcal{E}}$, $f'(\underline{\mathcal{U}})$ in (11) can be computed as

$$f'(\underline{\mathcal{U}}) = \sum_{\underline{e} \in \underline{\mathcal{U}}} Q_{\underline{e}} + \min_{\underline{\mathcal{R}} \subseteq \underline{\mathcal{U}}} \left\{ f(\underline{\mathcal{R}}) - \sum_{\underline{e} \in \underline{\mathcal{R}}} Q_{\underline{e}} \right\}, \ \forall \, \underline{\mathcal{U}} \subseteq \underline{\mathcal{E}}. \tag{18}$$

Then, since the function $f(\underline{\mathcal{R}}) - \sum_{\underline{e} \in \underline{\mathcal{R}}} Q_{\underline{e}}, \ \forall \ \underline{\mathcal{R}} \subseteq \underline{\mathcal{E}}$ is a submodular set function, we can solve the minimization in (18) using submodular function minimization routines that have a complexity polynomial in $|\underline{\mathcal{U}}|$ [30, 31]. Thus, from (12) we can conclude that $h(\underline{\mathcal{U}})$ can indeed be determined with a complexity polynomial in $|\underline{\mathcal{U}}|$.

We now propose simple observations that can considerably speed up the greedy algorithm

- *Lazy evaluations.* An important feature that speeds up the greedy algorithm substantially has been

[2]This assumption results in no loss of generality since the worst-case cost of computing $f(\underline{\mathcal{U}})$ is $O(NK^3)$.

discovered and exploited in [32,33]. In particular, due to the submodularity of the objective function the incremental gain offered by an element over any selected subset of elements not including it decreases monotonically as the selected subset grows larger. Thus, at any step in the algorithm, given a set of selected elements $\underline{\mathcal{S}}$ and an element $\underline{e} \in \underline{\mathcal{E}} \setminus \underline{\mathcal{S}}$ for which $h(\underline{\mathcal{S}} \cup \underline{e})$ has been evaluated, we do not have to evaluate $h(\underline{\mathcal{S}} \cup \underline{e}')$ for another element $\underline{e}' \in \underline{\mathcal{E}} \setminus \underline{\mathcal{S}}$, if we can assert that $h(\underline{\mathcal{S}} \cup \underline{e}) - h(\underline{\mathcal{S}}) \geq h(\underline{\mathcal{S}}' \cup \underline{e}') - h(\underline{\mathcal{S}}')$ where $\underline{\mathcal{S}}' \subseteq \underline{\mathcal{S}}$ denotes the set of selected elements at a previous step. This results in no loss of optimality with respect to the original greedy algorithm.

- *Exploiting subadditivity.* Suppose that at any step of the greedy algorithm we have a set of selected elements $\underline{\mathcal{S}}$. Further, let $\underline{e}_1 = (u, \mathbf{W}, \mathbf{c}_1)$ and $\underline{e}_2 = (u, \mathbf{W}, \mathbf{c}_2)$ be two elements in $\underline{\mathcal{E}} \setminus \underline{\mathcal{S}}$ such that $\mathbf{c}_1$ and $\mathbf{c}_2$ comprise of only one chunk each and are mutually non-intersecting. Then, letting $\underline{e}' = (u, \mathbf{W}, \mathbf{c}_1 + \mathbf{c}_2)$, we see that

$$h(\underline{\mathcal{S}} \cup \underline{e}') \leq h(\underline{\mathcal{S}} \cup \underline{e}_1 \cup \underline{e}_2) \leq h(\underline{\mathcal{S}} \cup \underline{e}_1) + h(\underline{\mathcal{S}} \cup \underline{e}_2) \tag{19}$$

where the first inequality stems from the fact that $h(\underline{\mathcal{S}} \cup \underline{e}')$ is monotonically increasing in the transmit PSD of $\underline{e}'$ and the second inequality stems from the monotonicity and subadditivity of $h(.)$. Thus, we have that

$$h(\underline{\mathcal{S}} \cup \underline{e}') \leq 2 \max\{h(\underline{\mathcal{S}} \cup \underline{e}_1), h(\underline{\mathcal{S}} \cup \underline{e}_2)\}. \tag{20}$$

Then if $\underline{\mathcal{S}} \cup \underline{e}_1, \underline{\mathcal{S}} \cup \underline{e}_2$ as well as $\underline{\mathcal{S}} \cup \underline{e}'$ satisfy all the constraints, we can evaluate $h(\underline{\mathcal{S}} \cup \underline{e}_1), h(\underline{\mathcal{S}} \cup \underline{e}_2)$ and skip evaluating $h(\underline{\mathcal{S}} \cup \underline{e}')$. By adopting this procedure over all elements in $\underline{\mathcal{E}} \setminus \underline{\mathcal{S}}$, we can ensure that the element selected will offer at-least $1/2$ the gain yielded by the locally optimal element. Then, using a well known result on the greedy algorithm with an approximately optimal selection at each step [27,28] we can conclude that this variation of our greedy algorithm will yield an approximation guarantee of $\frac{1/2}{1/2+M+1}$ when Assumptions 1 and 2 are satisfied.

3 Practical Modulation and Coding Schemes

In the LTE-A uplink a scheduled user can be assigned one out of three modulations (4, 16 & 64 QAM) and an outer Turbo-code whose coding rate is one out of several available choices. Since the available outer

codes are powerful and since the BS can employ near-optimal receivers (such as Turbo SIC) a reasonable choice for the achievable rate region is the following. Let $\mathcal{S}_{\underline{e}}$ denote the constellation (with unit average energy and cardinality $S_{\underline{e}}$) associated with element $\underline{e} \in \underline{\mathcal{E}}$. For any subset $\underline{\mathcal{A}} \subseteq \underline{\mathcal{E}}$ and any $n : 1 \leq n \leq N$, let $\mathcal{I}^{(n)}(\underline{\mathcal{A}})$ denote the mutual information evaluated for a point-to-point MIMO channel whose output can be modeled as

$$\mathbf{y}^{(n)} = \sum_{\underline{e} \in \underline{\mathcal{A}}} \sqrt{p_{\underline{e}}} c_{\underline{e}}(n) \mathbf{H}_{\underline{e}}^{(n)} \mathbf{W}_{\underline{e}}^{(n)} \mathbf{x}_{\underline{e}}^{(n)} + \mathbf{v}^{(n)}, \tag{21}$$

where $\mathbf{v}^{(n)} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ is the additive Gaussian noise and $\mathbf{x}_{\underline{e}}^{(n)} \in \mathcal{S}_{\underline{e}}^{N_t}$ is the input vector corresponding to element $\underline{e}$ whose entries are independently and uniformly drawn from $\mathcal{S}_{\underline{e}}$ and where $\mathbf{x}_{\underline{e}}^{(n)}, \mathbf{x}_{\underline{e}'}^{(n)}$ are mutually independent for any $\underline{e} \neq \underline{e}'$. Then, for any $\underline{\mathcal{U}} \subseteq \underline{\mathcal{E}}$ an achievable rate region is given by

$$\left\{ \mathbf{r} = [r_{\underline{e}}]_{\underline{e} \in \underline{\mathcal{U}}} \in {\rm I\!R}_+^{|\underline{\mathcal{U}}|} : \sum_{\underline{e} \in \underline{\mathcal{A}}} r_{\underline{e}} \leq \sum_{n=1}^{N} \mathcal{I}^{(n)}(\underline{\mathcal{A}}), \ \forall\, \underline{\mathcal{A}} \subseteq \underline{\mathcal{U}} \right\}. \tag{22}$$

Notice that in deriving (22) we have assumed an ideal BS receiver as well as no DFT spreading by each user, both of which allow for higher achievable rates.[3] Unfortunately, no closed form expressions are available for $\mathcal{I}^{(n)}(\underline{\mathcal{A}})$ and the rate region in (22) does not have a useful structure. Clearly the region defined before in (5) assuming Gaussian inputs is an outer bound which however can be loose. Here we obtain a tighter outer bound that also has a useful structure. We first offer the following result.

Proposition 1. *For any subset $\underline{\mathcal{A}} \subseteq \underline{\mathcal{E}}$ and any $n : 1 \leq n \leq N$, we have that*

$$\mathcal{I}^{(n)}(\underline{\mathcal{A}}) \leq \underbrace{\min_{\underline{\mathcal{R}} \subseteq \underline{\mathcal{A}}} \left\{ \log \left| \mathbf{I} + \sum_{\underline{e} \in \underline{\mathcal{A}} \setminus \underline{\mathcal{R}}} p_{\underline{e}} c_{\underline{e}}(n) \mathbf{H}_{\underline{e}}^{(n)} \mathbf{W}_{\underline{e}} (\mathbf{H}_{\underline{e}}^{(n)} \mathbf{W}_{\underline{e}})^\dagger \right| + \sum_{\underline{e} \in \underline{\mathcal{R}}} N_t \log(S_{\underline{e}}) \right\}}_{\triangleq g^{(n)}(\underline{\mathcal{A}})} \tag{23}$$

Further the set function $g : 2^{\underline{\mathcal{E}}} \to {\rm I\!R}_+$ defined as $g(\underline{\mathcal{A}}) = \sum_{n=1}^{N} g^{(n)}(\underline{\mathcal{A}}),\ \forall\, \underline{\mathcal{A}} \subseteq \underline{\mathcal{E}}$, is a rank function.

Proof. Consider any $\underline{\mathcal{A}} \subseteq \underline{\mathcal{E}}, n : 1 \leq n \leq N$ and the model in (21). Using the chain rule for mutual information along with the fact that the inputs corresponding to any two distinct elements of $\underline{\mathcal{A}}$ are

[3] Neglecting the per-user DFT spreading expands the rate region since the noise at the BS is assumed to be Gaussian and independent across RBs.

mutually independent, we can upper bound $\mathcal{I}^{(n)}(\underline{\mathcal{A}})$ as

$$\mathcal{I}^{(n)}(\underline{\mathcal{A}}) \leq \mathcal{I}^{(n)}(\underline{\mathcal{A}} \setminus \underline{\mathcal{R}}) + \sum_{\underline{e} \in \underline{\mathcal{R}}} \mathcal{I}^{(n)}(\underline{e}),$$

for any $\underline{\mathcal{R}} \subset \underline{\mathcal{A}}$. Since the cardinality of the input corresponding to element $\underline{e}$ is $S_{\underline{e}}^{N_t}$ we have that $\mathcal{I}^{(n)}(\underline{e}) \leq N_t \log(S_{\underline{e}})$. Then using the fact that for any given input covariance, Gaussian inputs (with the same covariance) maximize the mutual information (over the Gaussian noise channel model in (21)), we have that

$$\mathcal{I}^{(n)}(\underline{\mathcal{A}} \setminus \underline{\mathcal{R}}) \leq \log \left| \mathbf{I} + \sum_{\underline{e} \in \underline{\mathcal{A}} \setminus \underline{\mathcal{R}}} p_{\underline{e}} c_{\underline{e}}(n) \mathbf{H}_{\underline{e}}^{(n)} \mathbf{W}_{\underline{e}} (\mathbf{H}_{\underline{e}}^{(n)} \mathbf{W}_{\underline{e}})^{\dagger} \right|.$$

Since these arguments are valid for any subset $\underline{\mathcal{R}} \subseteq \underline{\mathcal{A}}$, we can deduce that (23) is true. The remaining result follows from basic definitions. $\square$

In this context, we note that the bound in (23) is a non-trivial generalization of a bound on the finite alphabet mutual information over a point-to-point fading channel employed in [34] to derive a tight lower bound on the outage probability. However, that bound when applied to our case would only yield $\mathcal{I}^{(n)}(\underline{e}) \leq \min\{\log |\mathbf{I} + p_{\underline{e}} c_{\underline{e}}(n) \mathbf{H}_{\underline{e}}^{(n)} \mathbf{W}_{\underline{e}} (\mathbf{H}_{\underline{e}}^{(n)} \mathbf{W}_{\underline{e}})^{\dagger}|, N_t \log(S_{\underline{e}})\}$ for any $\underline{e} \in \underline{\mathcal{E}}$.

Next, we outer bound the region in (22) as

$$\underline{\mathcal{T}}(\underline{\mathcal{U}}, g) \triangleq \left\{ \mathbf{r} = [r_{\underline{e}}]_{\underline{e} \in \underline{\mathcal{U}}} \in \mathbb{R}_+^{|\underline{\mathcal{U}}|} : \sum_{\underline{e} \in \underline{\mathcal{A}}} r_{\underline{e}} \leq g(\underline{\mathcal{A}}), \; \forall \; \underline{\mathcal{A}} \subseteq \underline{\mathcal{U}} \right\}. \tag{24}$$

Invoking Proposition 1 we use the fact that $g(.)$ is a rank function from which it follows that the region $\underline{\mathcal{T}}(\underline{\mathcal{U}}, g)$ is a polymatroid. Then invoking Lemma 2 we can infer the following result.

Proposition 2. *For any choice of selected elements $\underline{\mathcal{U}} \subseteq \underline{\mathcal{E}}$, the rate region $\underline{\mathcal{T}}(\underline{\mathcal{U}}, g') \triangleq \underline{\mathcal{T}}(\underline{\mathcal{U}}, g) \cap \underline{\mathcal{B}}(\underline{\mathcal{U}})$ is a polymatroid which is characterized by the rank function*

$$g'(\underline{\mathcal{A}}) = \min_{\underline{\mathcal{R}} \subseteq \underline{\mathcal{A}}} \left\{ g(\underline{\mathcal{A}} \setminus \underline{\mathcal{R}}) + \sum_{\underline{e} \in \underline{\mathcal{R}}} Q_{\underline{e}} \right\}, \; \forall \; \underline{\mathcal{A}} \subseteq \underline{\mathcal{U}}. \tag{25}$$

Then, upon by defining

$$h'(\underline{\mathcal{U}}) = \max_{\substack{\mathbf{r}=[r_{\underline{e}}]_{\underline{e}\in\underline{\mathcal{U}}} \\ \mathbf{r}\in\underline{\mathcal{T}}(\underline{\mathcal{U}},g')}} \left\{ \sum_{\underline{e}\in\underline{\mathcal{U}}} \alpha_{\underline{e}} r_{\underline{e}} \right\}, \quad \forall\, \underline{\mathcal{U}} \subseteq \underline{\mathcal{E}},$$

we consider the optimization problem

$$\begin{aligned} \max_{\underline{\mathcal{U}}\subseteq\underline{\mathcal{E}}} \{h'(\underline{\mathcal{U}})\} \ \text{ s.t.} \\ \underline{\mathcal{U}} \in \underline{\mathcal{I}}; \\ \mathbf{A}_I \mathbf{x}_{\underline{\mathcal{U}}} \leq \mathbf{1}_M; \ \ \mathbf{A}_C \mathbf{x}_{\underline{\mathcal{U}}} \leq \mathbf{1}_L. \end{aligned} \tag{26}$$

As before, it can be shown that the optimization problem in (26) is the maximization of a monotonic submodular function subject to one matroid and multiple knapsack constraints. Algorithm I and its associated results are thus applicable.

4 Simulation Results

In this section we present our simulation results. We simulate an uplink with 10 users, wherein the BS is equipped with four receive antennas. The system has 280 sub-carriers divided into 20 RBs (of size 14 sub-carriers each) available as data subcarriers that are used for serving the users. We assume 10 active users all of whom have identical maximum transmit powers. We use the SCM urban macro channel model (with co-polarized antennas having $10\lambda,\ 1\lambda$ separation at the BS and the mobile (user), respectively, and 15^o BS mean angular spread) to generate the channel between each user and the base-station. In all the results given below we assume an infinitely backlogged traffic model.

In Fig. 2, we assume that each user is equipped with two transmit antennas and can use an antenna selection codebook, i.e., $\mathcal{W} = \{[1;0],[0;1]\}$. The BS employs the optimal receiver and each user can employ an unconstrained (Gaussian) input alphabet. For simplicity, we assume no interference limit constraints and consider one control channel overhead constraint which imposes that no more than seven users can be scheduled. We plot the average cell spectral efficiency curves obtained when Algorithm I is employed by the BS scheduler with and without the control channel overhead constraint (denoted respectively by Algo-I-limit-AS and Algo-I-AS). Also plotted are the corresponding spectral efficiency curves obtained when

each user has only one transmit antenna (denoted respectively by Algo-I-limit and Algo-I). For each curve, we plot a corresponding upper bound by specializing a data-dependent upper bound from [32] which is applicable to any sub-modular function maximization (see also [33]). From the figure we observe that with and without antenna selection, the performance of Algorithm I is within $68 - 75\%$ of the data-dependent upper bound, which is superior to the worst case guarantee 1/2 (obtained by specializing the result in Theorem 3).

5 Conclusions

We considered resource allocation in the 3GPP LTE-A cellular uplink which allows for MIMO transmission from each scheduled user as well as multi-user scheduling wherein multiple users can be assigned the same time-frequency resource. We showed that the resulting resource allocation problem is NP-hard and then proposed constant-factor polynomial-time approximation algorithms.

References

[1] 3GPP, "TSG-RAN EUTRA, rel.10," *TR 36.213*, Dec 2010.

[2] W. Yu and W. Rhee, "Degrees of freedom in wireless multiuser spatial multiplex systems with multiple antennas," *IEEE Trans. Commun.*, vol. 54, pp. 1747–1753, Oct 2006.

[3] D. Tse and S. Hanly, "Multiaccess fading channels-part I: Polymatroid structure, optimal resource allocation, and throughput capacities," *IEEE Trans. Inform. Theory*, 1998.

[4] 3GPP, "TSG-RAN EUTRA, rel.8," *TR 36.101*, Jun 2011.

[5] C. Chekuri, J. Vondrak, and R. Zenklusen, "Dependent randomized rounding via exchange properties of combinatorial structures," *FOCS*, 2010.

[6] N. Bansal, N. Korula, V. Nagarajan, and A. Srinivasan, "On k-column sparse packing programs," *Proc. International Conference on Integer Programming and Combinatorial Optimization*, 2010.

[7] W. Yu and R. Liu, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *IEEE Trans. Commun.*, vol. 54, pp. 1310–1322, July 2006.

[8] W. Yu, R. Liu, and R. Cendrillon, "Dual optimization methods for multiuser orthogonal frequency division multiplex systems," *Proc. IEEE Global Telecommun. Conf. (Globecom)*, 2004.

[9] C. Wong, R. Cheng, K. Letaief, and R. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 1747 – 1758, Oct. 1999.

[10] S. Lee, S. Choudhury, A. Khoshnevis, S. Xu, and S. Lu, "Downlink MIMO with frequency-domain packet scheduling for 3GPP LTE," *IEEE Infocom*, 2009.

[11] A. Abrardo, P. Detti, and M. Moretti, "Message passing resource allocation for the uplink of multi-carrier systems," *Proc. IEEE Int. Conf. on Commun. (ICC)*, 2009.

[12] W. Noh, "A distributed resource control for fairness in ofdma systems: English-auction game with imperfect information," *Proc. IEEE Global Telecommun. Conf. (Globecom)*, 2008.

[13] K. Kim and Y. Han, "Joint subcarrier and power allocation in uplink OFDMA systems," *IEEE Commun. Let.*, vol. 9, pp. 526–528, Jan. 2006.

[14] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Trans. on Information Theory*, 2010.

[15] N. Ravindran and N. Jindal, "Limited feedback-based block diagonalization for the MIMO broadcast channel," *IEEE Journal Sel. Areas in Communications*, 2009.

[16] H. Zhang, N. Prasad, and S. Rangarajan, "MIMO downlink scheduling in LTE and LTE-advanced sustems," in *Proc. 2012 IEEE INFOCOM Miniconference*, (Orlando, FL), Mar. 2012.

[17] M. Andrews and L. Zhang, "Multiserver scheduling with contiguity constraints," *Proc. IEEE Infocom*, 2009.

[18] H. Yang, F. Ren, C. Lin, and J. Zhang, "Frequency-domain packet scheduling for 3GPP LTE uplink," *IEEE Infocom*, 2010.

[19] N. Prasad, H. Zhang, M. Jiang, G. Yue, and S. Rangarajan, "Resource allocation in 4G MIMO cellular uplink," *IEEE Globecom*, 2011.

[20] N. Prasad, H. Zhang, H. Zhu, and S. Rangarajan, "Multi-user scheduling in the 3GPP LTE cellular uplink," *Tech. Report: uploaded in arXiv cs.NI*, Jan. 2012.

[21] Y. Liu and E. Knightly, "Opportunistic fair scheduling over multiple wireless channels," in *Proc. 2003 IEEE INFOCOM*, (San Francisco, CA), Mar. 2003.

[22] H. Zhu, N. Prasad, and S. Rangarajan, "Precoder design for physical layer multicasting," *Tech. Report: uploaded in arXiv cs.NI, to appear ICC 2012*, Sept. 2011.

[23] S. Hanly and D. Tse, "Multi-access fading channels: Part II: Delay-limited capacities," *IEEE Trans. on Information Theory*, 1998.

[24] J. Edmonds, "Submodular functions, matroid, and certain polyhedra," *Combinatorial Structures and Their Applications*, 1970.

[25] C. Ko, J. Lee, and M. Queyranne, "An exact algorithm for maximum entropy sampling," *Operations Research*, 1995.

[26] J. O. Cerdeiral and P. Barcia, "When is a 0-1 knapsack a matroid?," *Portugaliae Mathematica*, 1995.

[27] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, "An analysis of approximations for maximizing submodular set functions," *Math. Programming*, 1978.

[28] G. L. Nemhauser and L. A. Wolsey, "Best algorithms for approximating the maximum of a submodular set function," *Math. Operations Research*, 1978.

[29] S. Sanayei and A. Nosratinia, "Capacity of MIMO channels with antenna selection," *IEEE Trans. on Information Theory*, 2007.

[30] S. Iwata, L. Fleischer, and S. Fujishige, "A combinatorial strongly polynomial algorithm for minimizing submodular functions," *Journal of ACM*, 2001.

[31] S. Iwata and J. B. Orlin, "A simple combinatorial algorithm for submodular function minimization," *Proc. ACM-SIAM Symposium on Discrete Algorithms*, 2009.

[32] M. Minoux, "Accelerated greedy algorithms for maximizing submodular set functions," *Optimization Techniques, LNCS*, 1978.

[33] A. Krause, J. Leskovec, C. Guestrin, J. VanBriesen, and C. Faloutsos, "Efficient sensor placement optimization for securing large water distribution networks," *Journal of Water Resources Planning and Management*, 2008.

[34] K. D. Nguyen, A. G. Fabregas, and L. K. Rasmussen, “A tight lower bound to the outage probability of discrete-input block-fading channels,” *IEEE Trans. on Information Theory*, 2007.

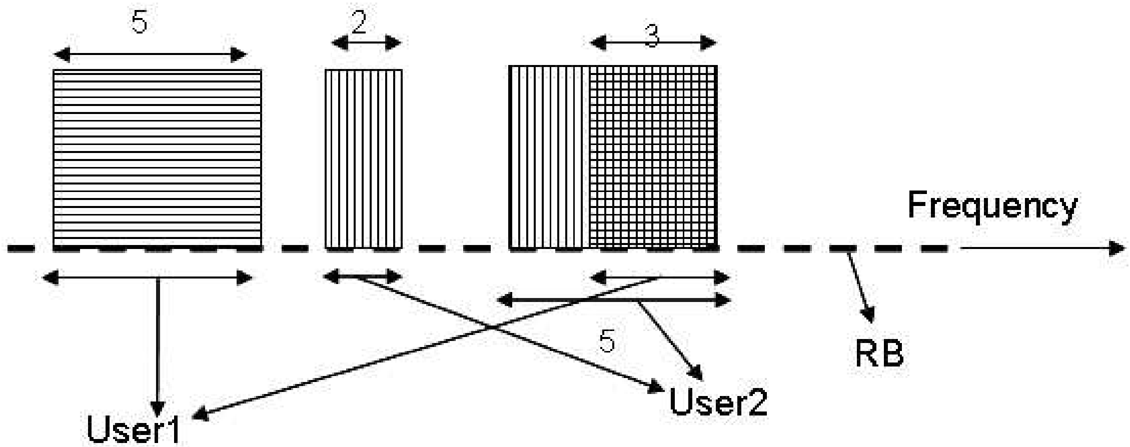

Figure 1: A Feasible RB Allocation in the LTE-A UL

Table 1: **Algorithm I: Greedy Algorithm for LTE-A UL MU-MIMO**

1: Initialize $\underline{\mathcal{S}} = \phi$
2: **Repeat**
3: Determine

$$\hat{\underline{e}} = \arg \max_{\substack{\underline{e} \in \underline{\mathcal{E}} \setminus \underline{\mathcal{S}} \\ \underline{\mathcal{S}} \cup \underline{e} \in \underline{\mathcal{I}}; \mathbf{A}_I \mathbf{x}_{\underline{\mathcal{S}} \cup \underline{e}} \leq \mathbf{1}_M; \mathbf{A}_C \mathbf{x}_{\underline{\mathcal{S}} \cup \underline{e}} \leq \mathbf{1}_L}} \{h(\underline{\mathcal{S}} \cup \underline{e})\} \tag{27}$$

and set $\hat{v} = h(\underline{\mathcal{S}} \cup \hat{\underline{e}}) - h(\underline{\mathcal{S}})$.
4: **If** $\hat{v} > 0$ **Then**
5: $\underline{\mathcal{S}} \leftarrow \underline{\mathcal{S}} \cup \underline{e}$
6: **End If**
7: **Until** $\hat{v} \leq 0$ or $\hat{\underline{e}} = \phi$
8: Output $\underline{\mathcal{S}}$.

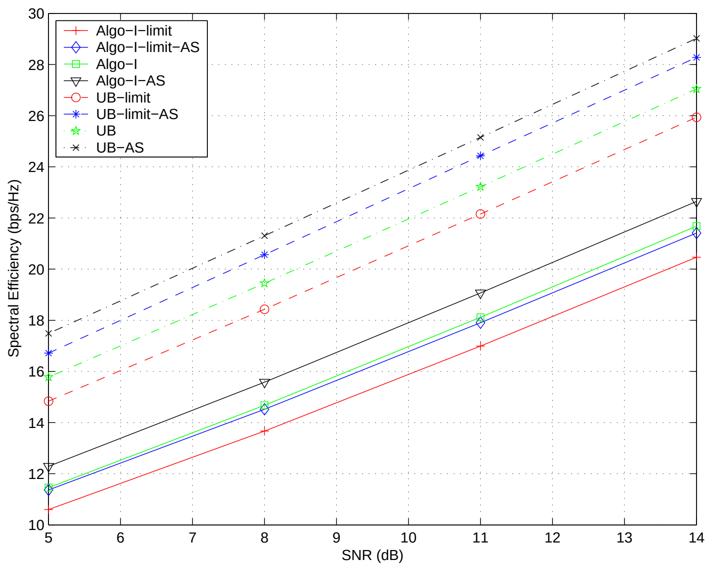

Figure 2: Average spectral efficiency versus SNR (dB): LTE-A MU-MIMO Scheduling.